

Impact of memory on clustering in spontaneous particle aggregation

Radek Erban

Mathematical Institute, University of Oxford,
 Radcliffe Observatory Quarter, Woodstock Road,
 Oxford, OX2 6GG, United Kingdom
radek.erban@maths.ox.ac.uk

Jan Haskovec

Mathematical and Computer Sciences and Engineering Division,
 King Abdullah University of Science and Technology,
 Thuwal 23955-6900, Kingdom of Saudi Arabia
jan.haskovec@kaust.edu.sa

Abstract

The effect of short-term and long-term memory on spontaneous aggregation of organisms is investigated using a stochastic agent-based model. Each individual modulates the amplitude of its random motion according to the perceived local density of neighbors. Memory is introduced via a chain of K internal variables that allow agents to retain information about previously encountered densities. The parameter K controls the effective length of memory. A formal mean-field limit yields a macroscopic Fokker–Planck equation, which provides a continuum description of the system in the large-population limit. Steady states of this equation are characterized to interpret the emergence and morphology of clusters. Systematic stochastic simulations in one- and two-dimensional spatial domains reveal that short- or moderate-term memory promotes coarsening, resulting in a smaller number of larger clusters, whereas long-term memory inhibits aggregation and increases the proportion of isolated individuals. Statistical analysis demonstrates that extended memory reduces the agents’ responsiveness to environmental stimuli, explaining the transition from aggregation to dispersion as K increases. These findings identify memory as a key factor controlling the collective organization of self-driven agents and provide a bridge between individual-level dynamics and emergent spatial patterns.

Relevance to Life Sciences. The presented framework provides a quantitative approach for studying how biological memory influences collective organization. Many biological systems, including bacterial populations, cellular assemblies, and insect groups, display aggregation behavior that depends on how past environmental information is integrated into current decision-making. The model bridges individual cognition (through memory) and population-level spatial structure (through clustering). It demonstrates that limited memory enhances aggregation by maintaining sensitivity to current local conditions, while excessively long memory diminishes responsiveness and disrupts coordination. This mechanism offers a potential explanation for experimentally observed adaptation timescales in biochemical and behavioral systems, such as the short-term biochemical memory of bacterial chemotaxis or transient social memory in gregarious insects. The results suggest that the duration of memory may have evolved to balance responsiveness and stability in collective behaviors, leading to efficient spatial organization under fluctuating environmental conditions.

Mathematical Content. The model is formulated as a system of coupled stochastic differential equations for N interacting agents, each endowed with K internal variables forming a linear memory cascade. The memory kernel quantifies how past information influences particle motion and determines the effective memory timescale. The formal mean-field limit, derived under the molecular chaos assumption, yields a nonlinear and nonlocal Fokker–Planck equation for the joint density of spatial and internal states. Analytical characterization of its steady states identifies conditions for spatially inhomogeneous equilibria corresponding to particle clusters. Numerical investigations employ direct stochastic simulations combined with density-based spatial clustering analysis to quantify aggregation statistics and elucidate how memory influences agents’ responsiveness to environmental stimuli. The study thus links microscopic stochastic processes with macroscopic pattern formation through a unified mathematical framework.

Keywords. Spontaneous aggregation; Stochastic particle systems; Memory effects; Fokker–Planck equation; Collective behaviour.

1 Introduction

Models with memory and delays are ubiquitous in explaining collective behaviour of organisms, ranging in scale from bacterial and amoeboid chemotaxis [3, 29] to behaviour of social insects and flocks of birds [53, 54]. In the case of bacteria, the memory is incorporated in purely biochemical terms, in the form of the signal transduction network capable of excitation and adaptation dynamics [52, 2]. This enables a cell to ‘memorize’ past environmental signals and compare to their current state to inform decision making [4]. In individual-based models, such memory effects can be accounted for by ordinary differential equation (ODE) or stochastic differential equation (SDE) models describing internal variables of each individual [27, 29]. A typical example of such internal variables are concentrations of key intracellular biochemical species. At the collective (macroscopic) level, populations of bacteria or other cells can be described by partial differential equations (PDEs) describing the density of cells and key environmental signals [28, 24], where memory can also be interpreted as delays in the processing of some signals. Incorporating such delays into mathematical models of collective phenomena can provide more accurate explanations of behavioural properties of groups of animals [55, 26], robots [56], or spreading processes in social and biological systems [49]. Moreover, non-Markovian characteristics resulting from ordering of interactions in temporal complex networks were identified as an important mechanism that alters causality and affects dynamical processes in social, technological and biological systems [50, 60]. Considering biological agents (cells or animals), a trade off between having no memory at all and remembering too much (including very distant and irrelevant past states) has been achieved by evolution. For example, bacterial ‘biochemical memory’ of past environmental signals evolved to last for the duration of a few seconds [4].

In this paper we focus on a model of spontaneous aggregation of animals or cells resulting from random diffusive motion of individuals [8, 36]. The individuals respond to the local population density observed in their neighbourhood by increasing or decreasing the amplitude of their random motion. This kind of behaviour has been observed in insects, for instance, the pre-social German cockroach (*B. germanica*), which are known to be attracted to dark, warm and humid places. However, it has been shown that cockroach larvae also aggregate spontaneously [40], i.e., in the absence of any environmental template or heterogeneity. Moreover, this type of dynamics can also be used to describe formation of P Granules in *C. elegans* embryos resulting in spontaneous protein aggregation [7]. We note that spontaneous aggregates form at random locations in space. If such spontaneous aggregates are to form on a regular lattice, additional interactions between individuals

must be incorporated into the model [44].

To incorporate memory into the first-order model of spontaneous aggregation [8], we introduce a set of $K \geq 1$ internal variables of each agent. Such internal variables can have different physical meanings in different biological models. In applications to bacterial chemotaxis, individuals compare current and past concentrations of extracellular signals, and internal variables can be identified with the concentration of molecules transducing the signal in detailed models of the chemotactic network [12]. The internal variables can also be viewed as effectively storing the memory of the state of the external environment in the vicinity of the individual. In our case, we consider model of the spontaneous aggregation (i.e., with no external chemical signal) and $K \geq 1$ internal variables store a time-dependent memory of how crowded the neighborhood was in the past. In particular, each internal variable describes a ‘layer’ of memory, which is subject to two effects: (i) production (or excitation) from the internal variable of the next higher order, and (ii) spontaneous decay with a constant rate. The internal variable of the highest (i.e., K -th) order is then subject to random excitation, with amplitude modulated by a nonlinear function of the number density of agents observed in the physical neighbourhood. We note that in the first-order model introduced in [8], the positions of the agents in the physical space were directly subjected to random excitation (Brownian motion). In particular, our investigation extends their model by introducing a chain of K internal variables that allow the agents to ‘remember’ the densities they encountered in the past.

Our results show that the introduction of memory with a few layers, K , leads to the formation of a smaller number of larger clusters of agents. This trend is observed until $K = 3$ or $K = 4$, depending on the spatial dimension of the studied system, as shown in Sections 6 and 7. When the number of memory layers K is increased further, i.e., as the memory becomes ‘longer’, its effect starts to be disruptive. This is manifested by the increasing proportion of ‘outliers’, which are the agents that are not part of any cluster. We therefore conclude that short-term or medium-term memory enhances spontaneous aggregation, while long-term memory disrupts it. This observation is ecologically relevant because many organisms - from bacteria to insects to higher animals - do not respond to their environment instantaneously but instead use some form of memory or internal state to modulate their behavior. Species that benefit from aggregation (e.g., for warmth, protection, or reproduction) may have developed memory systems that retain only short-to-medium-term environmental information. Conversely, overly long memory may reduce responsiveness to current cues and hinder effective social or spatial organization. Therefore, organisms must balance between being too reactive (leading to unstable or erratic behavior) and being too inertial (failing to adapt to rapidly changing environments). This is consistent with ecological and evolutionary studies that view memory as an adaptive trait shaped by natural selection, see, e.g., [18, 14].

There is a vast body of literature focusing on mathematical models of aggregation dynamics and their analysis, with the number of papers counting in hundreds. These models include not only aggregation of agents or particles in the physical space (e.g., [40, 7, 44, 41, 37]), but also in abstract phase space, i.e., models of opinion formation [38, 39], flocking and swarming [57, 19, 48] or pattern formation resulting from attraction-repulsion interactions [20, 9]. For a relatively recent survey of models of collective motion we refer to [58]. Many variants of these models have been studied, including models with nonlocal interactions [46, 5, 10], topological interactions [34], hydrodynamic models [15], models with hierarchy [51, 59] and leaders [17], models with noise [11] and delay [16, 35] and combination of those [26]. In [1] the authors considered a variant of the Cucker-Smale flocking model with a Caputo fractional derivative in the time variable. The presence of the fractional derivative can be interpreted as a form of memory. The authors derived sufficient conditions for asymptotic flocking, which occurs at an algebraic rate, in contrast to the exponential flocking asymptotics in the original Cucker-Smale model [19]. Memory induced aggregation in collective

foraging was studied in [47]. Using an agent-based model, the authors argued that aggregation around salient patches can occur through formation of collective memory realized through local interactions and global displacement using Lévy walks. The impact of emergent memory on opinion dynamics was analyzed in [6], with the aim to explain how the exposure of the society to certain events deeply changes people's perception of the present and future. An analytical way was proposed to measure how much information a society can remember when an extensive number of news items was presented.

In all the aforementioned models, aggregation arises from explicit deterministic forces acting on the particles, typically expressed as a superposition of pairwise interactions. Eventual random effects only disrupt the aggregation. In contrast, in the spontaneous aggregation model studied in this paper, each particle is merely subject to Brownian motion with amplitude modulated by a nonlinear function of the density. In particular, on the microscopic level of description, there is no deterministic force acting on the particles and the interaction between them is indirect. In particular, the model cannot be formulated in terms of pairwise particle interactions. Therefore, aggregation in our model arises solely from a stochastic process and we cannot expect formation of regular patterns, like flocks, swarms or opinion clusters as in the aforementioned models.

The paper is organized as follows. In Section 2 we describe the (first-order) spontaneous aggregation model and summarize its main properties studied in the literature [8]. In Section 3 we introduce memory into this model and infer the main mathematical properties of the model with memory. In Section 4 we derive the formal macroscopic limit of the system as the number of agents tends to infinity, obtaining the corresponding Fokker-Planck equation. In Section 5 we then characterize its steady states to gain an insight into the patterns (clusters) formed by the system. In Section 6 we report the results of the stochastic simulations of the individual-based model in the spatially one-dimensional setting, while the two-dimensional results are presented in Section 7. Here we also provide a statistical evidence that the long-term memory inhibits the particles' responsivity to environmental stimuli. We conclude with the discussion in Section 8.

2 Spontaneous aggregation model without memory

The individual-based stochastic model introduced in reference [8] under the name ‘direct aggregation model’ consists of a group of $N \geq 2$ biological agents (cells or animals), characterized by their positions $\mathbf{x}_i(t) \in \mathbb{R}^d$, with spatial dimension $d \in \{1, 2, 3\}$ and $i \in [N]$, where we have denoted the set of indices by $[N] := \{1, 2, \dots, N\}$. Every individual senses the average density of its close neighbours, given by

$$\vartheta_i(t) = \frac{1}{N} \sum_{j \in [N]} W(\mathbf{x}_i(t) - \mathbf{x}_j(t)), \quad \text{for } i \in [N], \quad (2.1)$$

where $W(\mathbf{x}) = w(|\mathbf{x}|)$ with the weight function $w : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ assumed to be bounded, nonnegative, nonincreasing and integrable on \mathbb{R}^d . Without loss of generality we impose the normalization

$$\int_{\mathbb{R}^d} W(\mathbf{x}) d\mathbf{x} = 1. \quad (2.2)$$

A generic example of w is the (properly normalized) characteristic function of the interval $[0, R]$, corresponding to the sampling radius $R > 0$. The average density ϑ_i is then simply the fraction of individuals located within the distance R from the i -th individual. The individual positions are subject to a random walk with modulated amplitude, described by the system of coupled SDEs

$$d\mathbf{x}_i(t) = G(\vartheta_i) dB_i^t, \quad \text{for } i \in [N], \quad (2.3)$$

where \mathbf{B}_i^t are independent d -dimensional Brownian motions. The *response function* $G : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is assumed to be globally bounded, nonnegative and decreasing. The monotonicity of G is implied by the modeling assumption that the individuals respond to higher perceived population densities in their vicinity by reducing the amplitude of their random walk.

In reference [8], the mean-field limit (as $N \rightarrow \infty$) of the individual-based model (2.1)–(2.3) is studied, describing the system in terms of the number particle density $\varrho \equiv \varrho(t, \mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^d$. The time evolution of the density ϱ is subject to

$$\frac{\partial \varrho}{\partial t} = \frac{1}{2} \Delta (G(W * \varrho)^2 \varrho), \quad (2.4)$$

with the convolution given by $W * \varrho(\mathbf{x}) := \int_{\mathbb{R}^d} W(\mathbf{x} - \mathbf{z}) \varrho(\mathbf{z}) d\mathbf{z}$. In fact, it is more instructive to expand the derivative in equation (2.4) and formulate it as the convection-diffusion equation

$$\frac{\partial \varrho}{\partial t} = \frac{1}{2} \nabla \cdot (\varrho \nabla [G(W * \varrho)^2] + G(W * \varrho)^2 \nabla \varrho). \quad (2.5)$$

Here we identify the convection term $\nabla \cdot (\varrho \nabla [G(W * \varrho)^2])$ as the ‘driving force’ responsible for the eventual formation of aggregates (recall the monotonicity assumption on G). Then the following condition for aggregation, emanating from a perturbation of a given constant steady state $\varrho \equiv \varrho_0 > 0$, can be derived [8],

$$\operatorname{Re} \hat{W}(\boldsymbol{\xi}) > -\frac{G(\varrho_0)}{2 G'(\varrho_0) \varrho_0} \quad \text{for some } \boldsymbol{\xi} \in \mathbb{R}^d, \quad (2.6)$$

where $\hat{W} = \hat{W}(\boldsymbol{\xi})$ is the Fourier transform of W and $G'(\varrho_0)$ denotes the derivative of G at ϱ_0 . Note that due to the monotonicity assumption, we have $G'(\varrho_0) < 0$. For a fixed kernel $W = W(\mathbf{x})$ one can interpret the inequality (2.6) as a condition on the response function G to be steep enough in the neighbourhood of ϱ_0 . Then the effect of the convection term in equation (2.5) is stronger than the smoothing effect of the diffusion, and the resulting instability leads to formation of aggregates. These aggregates persist as a steady state solution in the large time limit.

3 Spontaneous aggregation model with memory

We now introduce memory into the spontaneous aggregation model (2.1)–(2.3) using a set of $K \in \mathbb{N}$ internal variables describing K memory ‘layers’. Each individual is then characterised by its position $\mathbf{x}_i = \mathbf{x}_i(t) \in \mathbb{R}^d$ and internal variables $\mathbf{y}_i^k = \mathbf{y}_i^k(t) \in \mathbb{R}^d$ for $i \in [N]$ and $k \in [K]$. Equation (2.3) is generalized for $i \in [N]$ to the following system of SDEs,

$$\begin{aligned} \frac{d\mathbf{x}_i(t)}{dt} &= \mathbf{y}_i^1(t), \\ \varepsilon_k \frac{d\mathbf{y}_i^k(t)}{dt} &= -\alpha_k \mathbf{y}_i^k(t) + \mathbf{y}_i^{k+1}(t), \quad \text{for } k \in [K-1], \\ \varepsilon_K d\mathbf{y}_i^K(t) &= -\alpha_K \mathbf{y}_i^K(t) dt + G(\vartheta_i(t)) d\mathbf{B}_i^t, \end{aligned} \quad (3.1)$$

where \mathbf{B}_i^t are again independent d -dimensional Brownian motions. The positive constants $\alpha_k > 0$, for $k \in [K]$, are the relaxation coefficients, and parameters $\varepsilon_k > 0$, for $k \in [K]$, define the time scales. The perceived densities $\vartheta_i = \vartheta_i(t)$ are given by (2.1).

The model with memory (3.1) has been introduced in a way that it can be reduced to the original spontaneous aggregation model (2.1)–(2.3) if we set $\alpha_k := 1$ for all $k \in [K]$ and pass to

the limit $\varepsilon_k \rightarrow 0$ for all $k \in [K]$. However, it can be also formulated in a more general way, which provides connections with the memory models in the literature [31, 47, 42]. Denoting the j -th component of $\mathbf{y}_i^k \in \mathbb{R}^d$ by $y_{i,j}^k$, for $i \in [N]$, $j \in [d]$ and $k \in [K]$, we introduce the following notation

$$\mathbf{y}_{i,j} = [y_{i,j}^1, y_{i,j}^2, \dots, y_{i,j}^K] \in \mathbb{R}^K \quad \text{for } i \in [N] \text{ and } j \in [d]. \quad (3.2)$$

To further analyze the behaviour of model (3.1), we fix

$$\varepsilon_k := 1 \quad \text{for all } k \in [K]. \quad (3.3)$$

Then the SDE system for $\mathbf{y}_{i,j} = \mathbf{y}_{i,j}(t) \in \mathbb{R}^K$, with $i \in [N]$ and $j \in [d]$, can be written in matrix form as

$$d\mathbf{y}_{i,j}(t) = A^{(K)} \mathbf{y}_{i,j}(t) dt + \boldsymbol{\beta}_i^{(K)}(t) dB_{i,j}^t \quad (3.4)$$

where $B_{i,j}^t$ is the j -th component of \mathbf{B}_i^t (i.e., the standard one-dimensional Brownian motion), and the constant matrix $A^{(K)} \in \mathbb{R}^{K \times K}$ and the vectors $\boldsymbol{\beta}_i^{(K)} = \boldsymbol{\beta}_i^{(K)}(t) \in \mathbb{R}^K$ are given by

$$A^{(K)} := \begin{pmatrix} -\alpha_1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -\alpha_2 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & 1 & 0 \\ 0 & & & -\alpha_{K-1} & 1 \\ 0 & \dots & 0 & -\alpha_K & \end{pmatrix} \quad \text{and} \quad \boldsymbol{\beta}_i^{(K)}(t) := \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ G(\vartheta_i(t)) \end{bmatrix}. \quad (3.5)$$

The reformulation of model (3.1) into the SDE form (3.4) aligns it more closely with the formulation of the reinforced diffusion model for memory-mediated animal movement [31]. The key difference in our study is that we focus on spontaneous aggregation, whereas some models in the literature examine aggregation in response to environmental cues to locate specific targets [47, 42]. Additionally, we explicitly describe the memory variables, whereas some models introduce memory implicitly through delays [31]. Model (3.4) can also be rewritten in the delay form, because the solution of the SDE system (3.4) is given by the variation-of-constants formula [45, Section 3.3]

$$\mathbf{y}_{i,j}(t) = \exp[tA^{(K)}] \mathbf{y}_{i,j}(0) + \int_0^t \exp[(t-s)A^{(K)}] \boldsymbol{\beta}_i^{(K)}(s) dB_{i,j}^s, \quad (3.6)$$

where $\mathbf{y}_{i,j}(0)$ is a prescribed initial condition at time $t = 0$. The SDE form of the memory model (3.4) can, in general, be used to describe the ‘chemical memory’ of unicellular organisms, where internal variables \mathbf{y}_i^k can be identified with concentrations of key chemical species of signal transduction networks [28, 29]. The behaviour of such memory models will depend on the properties of matrix $A^{(K)} \in \mathbb{R}^{K \times K}$, on its eigenvalues and eigenvectors. In this paper, we use the specific form of the general model (3.4) given by (3.5). In particular, choosing $\mathbf{y}_{i,j}(0) = 0$ in equation (3.6), we have for $y_{i,j}^1 \equiv y_{i,j}^1(t)$

$$y_{i,j}^1(t) = \int_0^t \kappa(t-s) G(\vartheta_i(s)) dB_{i,j}^s, \quad \text{with} \quad \kappa(t) := \left\{ \exp[tA^{(K)}] \right\}_{1,K},$$

where $\kappa(t)$ is the $(1, K)$ -th element of the matrix $\exp[tA^{(K)}] \in \mathbb{R}^{K \times K}$. To get further insight into the model behaviour we assume

$$\alpha_k := \alpha \quad \text{for all } k \in [K], \quad (3.7)$$

where $\alpha > 0$ is a single parameter of the model. Then the kernel $\kappa = \kappa(t)$ can be evaluated explicitly, because the matrix $A^{(K)}$ in (3.5) takes the form

$$A^{(K)} = -\alpha I + U,$$

with U the upper-diagonal matrix with ones on the first upper diagonal. In particular, all eigenvalues of matrix $A^{(K)}$ are equal to $-\alpha$. Since U commutes with the identity matrix, we have

$$\exp[tA^{(K)}] = \exp(-\alpha t) \exp[tU].$$

Moreover, U being upper diagonal, its K -th power U^K is the zero matrix, so that

$$\exp(tU) = \sum_{k=0}^{K-1} \frac{t^k U^k}{k!}.$$

Consequently, the $(1, K)$ -th element of $\exp[tU]$ equals to $t^{K-1}/(K-1)!$ and we have

$$\kappa(t) = \left\{ \exp[tA^{(K)}] \right\}_{1,K} = \frac{e^{-\alpha t} t^{K-1}}{(K-1)!}. \quad (3.8)$$

Using equation (3.1), we obtain

$$\frac{d\mathbf{x}_i(t)}{dt} = \frac{1}{(K-1)!} \int_0^t e^{-(t-s)\alpha} (t-s)^{K-1} G(\vartheta_i(s)) d\mathbf{B}_i^s, \quad (3.9)$$

with $\vartheta_i = \vartheta_i(s) \in \mathbb{R}$ given by equation (2.1). Let us note that this formulation allows for generalization of the model to noninteger values of $K > 0$ by writing the factorial $(K-1)!$ in terms of the Gamma-function as $\Gamma(K)$.

To understand how the choice of the number of layers K impacts the ‘length’ of the memory, we calculate the mean of the normalized version of the kernel $\kappa = \kappa(t)$,

$$\left(\int_0^{+\infty} \kappa(t) dt \right)^{-1} \left(\int_0^{+\infty} t \kappa(t) dt \right) = \frac{K}{\alpha}. \quad (3.10)$$

We observe that the length of the memory increases linearly with K . In Sections 6 and 7, we will numerically investigate the impact of the choice of K on the clustering properties of the individual-based model with memory given by equations (2.1) and (3.9). Let us note that the kernel (3.8) takes very different shapes even if the length (3.10) is kept constant. For instance, in Figure 1 we plot $\kappa = \kappa(t)$ for the length

$$\frac{K}{\alpha} = \left\{ \frac{1}{1}, \frac{2}{2}, \frac{3}{3} \right\} = 1 \text{ (left panel)}, \quad \text{and} \quad \frac{K}{\alpha} = \left\{ \frac{1}{0.5}, \frac{2}{1}, \frac{4}{2} \right\} = 2 \text{ (right panel)}.$$

In particular, for $K = 1$ the kernel reduces to $\kappa(t) = e^{-\alpha t}$, i.e., is globally strictly decreasing on $[0, +\infty)$, while for $K \geq 2$ it has a unique strict maximum in $(0, +\infty)$. Therefore, it is not sufficient to only explore the case $K = 1$ and tune the value of $\alpha > 0$ to modulate the length of the memory (3.10). In this paper we chose to fix the value $\alpha = 1$ and explore the impact of different choices of the number of layers K on the clustering properties of the system.

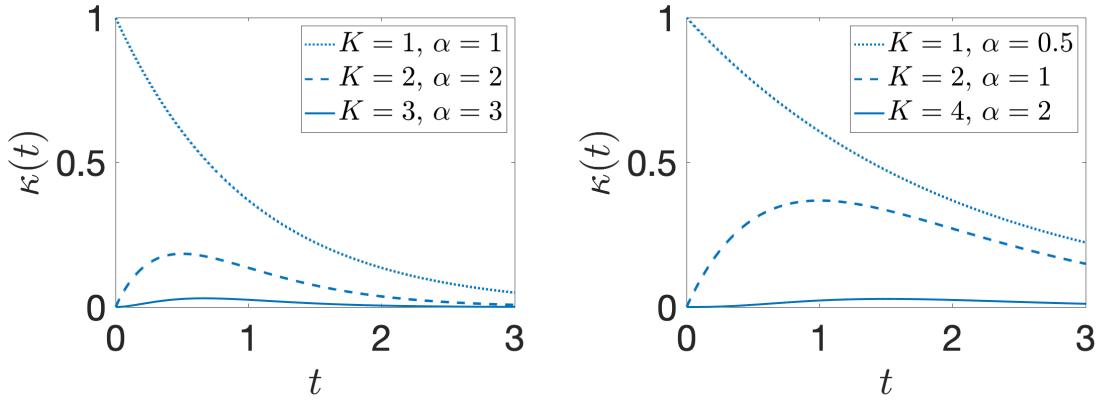


Figure 1: Plots of the kernel $\kappa = \kappa(t)$ given by (3.8) with $K/\alpha = 1$ (left panel) and with $K/\alpha = 2$ (right panel).

Finally, let us note that this individual-based model can be also viewed for $K = 1$ as a simplified version of what is called ‘the second-order model’ in reference [8]. Indeed, if $K = 1$, then the SDE system (3.1) reduces to

$$\begin{aligned} \frac{d\mathbf{x}_i(t)}{dt} &= \mathbf{y}_i^1(t), \\ d\mathbf{y}_i^1(t) &= -\alpha \mathbf{y}_i^1(t) dt + G(\vartheta_i(t)) dB_i^t. \end{aligned} \quad (3.11)$$

In particular, we can interpret the internal variable $\mathbf{y}_i^1 = \mathbf{y}_i^1(t)$ as the velocity of the i -th agent in the physical space. The velocity is then subject to linear damping with intensity $\alpha > 0$ and stochastic forcing with amplitude modulated by the term $G(\vartheta_i(t))$ with $\vartheta_i = \vartheta_i(t)$ given by (2.1).

4 Formal derivation of the Fokker-Planck equation

The formal large population limit as $N \rightarrow \infty$ of the system (3.1) with the average density given by (2.1) can be derived by a generalization of the procedure used in [8, Section 3]. Since some nontrivial steps were omitted in [8], we include the detailed formal derivation here. The main idea is to ‘linearize’ (3.1) by introducing the average densities $\vartheta_i = \vartheta_i(t)$ for $i \in [N]$ as a set of N independent variables. According to (2.1) and the first equation of (3.1), these new variables are subject to the system of differential equations

$$\frac{d\vartheta_i}{dt} = \frac{1}{N} \sum_{j \in [N]} \nabla W(\mathbf{x}_i - \mathbf{x}_j) \cdot (\mathbf{y}_i^1 - \mathbf{y}_j^1), \quad \text{for } i \in [N].$$

One then applies the Itô formula [45] to write down the Liouville equation [13] for the N -particle density $f^N = f^N(t, \mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{y}_1^1, \dots, \mathbf{y}_N^1, \dots, \mathbf{y}_1^K, \dots, \mathbf{y}_N^K, \vartheta_1, \dots, \vartheta_N)$,

$$\begin{aligned} \frac{\partial f^N}{\partial t} + \sum_{i \in [N]} \mathbf{y}_i^1 \cdot \nabla_{\mathbf{x}_i} f^N &+ \sum_{i \in [N]} \sum_{k \in [K-1]} \frac{1}{\varepsilon_k} \nabla_{\mathbf{y}_i^k} \cdot ((\mathbf{y}_i^{k+1} - \alpha_k \mathbf{y}_i^k) f^N) \\ &- \frac{1}{\varepsilon_K} \sum_{i \in [N]} \nabla_{\mathbf{y}_i^K} \cdot (\alpha_K \mathbf{y}_i^K f^N) \\ &+ \frac{1}{N} \sum_{i \in [N]} \sum_{j \in [N]} \nabla W(\mathbf{x}_i - \mathbf{x}_j) \cdot (\mathbf{y}_i^1 - \mathbf{y}_j^1) \frac{\partial f^N}{\partial \vartheta_i} \\ &= \frac{1}{2\varepsilon_K} \sum_{i \in [N]} \Delta_{\mathbf{y}_i^K} (G(\vartheta_i)^2 f^N). \end{aligned}$$

Integrating with respect to the variables $(\mathbf{x}_i, \mathbf{y}_i^1, \dots, \mathbf{y}_i^K, \vartheta_i)$ for $i = 2, \dots, N$ and adopting the usual molecular chaos assumption on vanishing particle correlations, we obtain the Vlasov equation for the one-particle marginal density $f^1 = f^1(t, \mathbf{x}, \mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^K, \vartheta)$. For brevity, we introduce the notation $\mathbf{y} := (\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^K)$. Then, the Vlasov equation for $f^1 = f^1(t, \mathbf{x}, \mathbf{y}, \vartheta)$ reads

$$\begin{aligned} \frac{\partial f^1}{\partial t} + \mathbf{y}^1 \cdot \nabla_{\mathbf{x}} f^1 &+ \sum_{k=1}^{K-1} \frac{1}{\varepsilon_k} \nabla_{\mathbf{y}^k} \cdot ((\mathbf{y}^{k+1} - \alpha_k \mathbf{y}^k) f^1) \\ &- \frac{1}{\varepsilon_K} \nabla_{\mathbf{y}^K} \cdot (\alpha_K \mathbf{y}^K f^1) + \frac{\partial}{\partial \vartheta} (F[f^1] f^1) \\ &= \frac{1}{2\varepsilon_K} \Delta_{\mathbf{y}^K} (G(\vartheta)^2 f^1), \end{aligned}$$

where we denoted

$$F[f^1](\mathbf{x}, \mathbf{y}^1) := \int_{\mathbb{R}^{d(K+1)+1}} \nabla W(\mathbf{x} - \mathbf{x}_*) \cdot (\mathbf{y}^1 - \mathbf{y}_*^1) f_*^1 d\mathbf{x}_* d\mathbf{y}_* d\vartheta_*, \quad (4.1)$$

with $f_*^1 := f^1(t, \mathbf{x}_*, \mathbf{y}_*, \vartheta_*)$. The weak formulation of the Vlasov equation with test function $\psi = \psi(\mathbf{x}, \mathbf{y}, \vartheta)$ reads

$$\begin{aligned} \frac{d}{dt} \int f^1 \psi &= \int (\mathbf{y}^1 f^1) \cdot \nabla_{\mathbf{x}} \psi + \sum_{k=1}^{K-1} \frac{1}{\varepsilon_k} \int ((\mathbf{y}^{k+1} - \alpha_k \mathbf{y}^k) f^1) \cdot \nabla_{\mathbf{y}^k} \psi \\ &\quad - \frac{1}{\varepsilon_K} \int (\alpha_K \mathbf{y}^K f^1) \cdot \nabla_{\mathbf{y}^K} \psi \\ &\quad + \int (F[f^1] f^1) \frac{\partial \psi}{\partial \vartheta} + \frac{1}{2\varepsilon_K} \int (G(\vartheta)^2 f^1) \Delta_{\mathbf{y}^K} \psi, \end{aligned}$$

where all integrals are with respect to the variables $(\mathbf{x}, \mathbf{y}, \vartheta) \in \mathbb{R}^{d(K+1)+1}$. We now introduce the ansatz

$$f^1(t, \mathbf{x}, \mathbf{y}, \vartheta) = f(t, \mathbf{x}, \mathbf{y}) \delta(\vartheta - W * \varrho) \quad (4.2)$$

with δ the Dirac delta function,

$$W * \varrho(t, \mathbf{x}) := \int_{\mathbb{R}^d} W(\mathbf{x} - \mathbf{x}_*) \varrho(t, \mathbf{x}_*) d\mathbf{x}_*$$

and the particle number density

$$\varrho(t, \mathbf{x}) := \int_{\mathbb{R}^{dK+1}} f^1(t, \mathbf{x}, \mathbf{y}, \vartheta) d\mathbf{y} d\vartheta. \quad (4.3)$$

Inserting (4.2) into the weak formulation of the Vlasov equation and denoting

$$\varphi(t, \mathbf{x}, \mathbf{y}) := \psi(\mathbf{x}, \mathbf{y}, W * \varrho(t, \mathbf{x}))$$

gives

$$\begin{aligned} \frac{d}{dt} \int_{\bullet} f \varphi &= \int (\mathbf{y}^1 f^1) \cdot \nabla_{\mathbf{x}} \psi + \sum_{k=1}^{K-1} \frac{1}{\varepsilon_k} \int_{\bullet} ((\mathbf{y}^{k+1} - \alpha_k \mathbf{y}^k) f) \cdot \nabla_{\mathbf{y}^k} \varphi \\ &\quad - \frac{1}{\varepsilon_K} \int_{\bullet} (\alpha_K \mathbf{y}^K f) \cdot \nabla_{\mathbf{y}^K} \varphi \\ &\quad + \int (F[f^1] f^1) \frac{\partial \psi}{\partial \vartheta} + \frac{1}{2\varepsilon_K} \int_{\bullet} (G(W * \varrho)^2 f) \Delta_{\mathbf{y}^K} \varphi, \end{aligned} \quad (4.4)$$

where here and in the sequel the symbol \int_{\bullet} denotes integration with respect to the variables $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{d(K+1)}$, while the plain integral symbol means integration with respect to $(\mathbf{x}, \mathbf{y}, \vartheta) \in \mathbb{R}^{d(K+1)+1}$. With the chain rule for the derivative we have

$$\nabla_{\mathbf{x}} \varphi = \nabla_{\mathbf{x}} \psi + \nabla_{\mathbf{x}} (W * \varrho) \frac{\partial \psi}{\partial \vartheta}.$$

Consequently, the first term on the left-hand side of (4.4) reads

$$\int (\mathbf{y}^1 f^1) \cdot \nabla_{\mathbf{x}} \psi = \int_{\bullet} (\mathbf{y}^1 f) \cdot \nabla_{\mathbf{x}} \varphi - \int (\mathbf{y}^1 f^1) \cdot \nabla_{\mathbf{x}} (W * \varrho) \frac{\partial \psi}{\partial \vartheta}. \quad (4.5)$$

Moreover, using (4.1), we write

$$\int (F[f^1] f^1) \frac{\partial \psi}{\partial \vartheta} = \iint \nabla W(\mathbf{x} - \mathbf{x}_*) \cdot \mathbf{y}^1 f_*^1 f^1 \frac{\partial \psi}{\partial \vartheta} - \iint \nabla W(\mathbf{x} - \mathbf{x}_*) \cdot \mathbf{y}_*^1 f_*^1 f^1 \frac{\partial \psi}{\partial \vartheta},$$

where the double integral sign \iint means integral with respect to the variables $(\mathbf{x}, \mathbf{y}, \vartheta, \mathbf{x}_*, \mathbf{y}_*, \vartheta_*) \in \mathbb{R}^{2d(K+1)+2}$. We now observe that the first term on the right-hand side

$$\iint \nabla W(\mathbf{x} - \mathbf{x}_*) \cdot \mathbf{y}^1 f_*^1 f^1 \frac{\partial \psi}{\partial \vartheta} = \int \nabla_{\mathbf{x}} (W * \varrho) \cdot \mathbf{y}^1 f^1 \frac{\partial \psi}{\partial \vartheta}$$

cancels the last term in (4.5). Moreover, the co-ordinate change $(\mathbf{x}, \mathbf{y}, \vartheta) \rightleftarrows (\mathbf{x}_*, \mathbf{y}_*, \vartheta_*)$, using the antisymmetry $\nabla W(\mathbf{x} - \mathbf{x}_*) = -\nabla W(\mathbf{x}_* - \mathbf{x})$, gives

$$\begin{aligned} - \iint \nabla W(\mathbf{x} - \mathbf{x}_*) \cdot \mathbf{y}_*^1 f_*^1 f^1 \frac{\partial \psi}{\partial \vartheta} &= \iint \nabla W(\mathbf{x} - \mathbf{x}_*) \cdot \mathbf{y}^1 f^1 f_*^1 \frac{\partial \psi_*}{\partial \vartheta} \\ &= \int_{\bullet} \nabla_{\mathbf{x}} \left[W \circledast \left(f \frac{\partial \psi}{\partial \vartheta} \right) \right] \cdot \mathbf{y}^1 f, \end{aligned}$$

where we denoted

$$W \circledast \left(f \frac{\partial \psi}{\partial \vartheta} \right) (t, \mathbf{x}) := \int_{\mathbb{R}^{d(K+1)}} W(\mathbf{x} - \mathbf{x}_*) f(t, \mathbf{x}_*, \mathbf{y}) \frac{\partial \psi}{\partial \vartheta}(\mathbf{x}_*, \mathbf{y}, W * \varrho(t, \mathbf{x})) d\mathbf{y} d\mathbf{x}_*.$$

We conclude that

$$\int (\mathbf{y}^1 f^1) \cdot \nabla_{\mathbf{x}} \psi + \int (F[f^1] f^1) \frac{\partial \psi}{\partial \vartheta} = \int_{\bullet} (\mathbf{y}^1 f) \cdot \nabla_{\mathbf{x}} \left[\varphi + W \circledast \left(f \frac{\partial \psi}{\partial \vartheta} \right) \right].$$

Finally, for the time derivative term in (4.4) we again use the chain rule

$$\frac{\partial \varphi}{\partial t} = \frac{\partial \psi}{\partial \vartheta} \left(W * \frac{\partial \varrho}{\partial t} \right)$$

and obtain

$$\begin{aligned} \frac{d}{dt} \int_{\bullet} f \varphi &= \int_{\bullet} \frac{\partial f}{\partial t} \varphi + \int_{\bullet} f \frac{\partial \psi}{\partial \vartheta} \left(W * \frac{\partial \varrho}{\partial t} \right) \\ &= \int_{\bullet} \frac{\partial f}{\partial t} \varphi + \int_{\bullet} \frac{\partial \varrho}{\partial t} W * \left(f \frac{\partial \psi}{\partial \vartheta} \right) \\ &= \int_{\bullet} \frac{\partial f}{\partial t} \left[\varphi + W \circledast \left(f \frac{\partial \psi}{\partial \vartheta} \right) \right]. \end{aligned}$$

Therefore, denoting $\xi := \varphi + W \circledast \left(f \frac{\partial \psi}{\partial \vartheta} \right)$ and noting that $\nabla_{\mathbf{y}} \xi = \nabla_{\mathbf{y}} \varphi$, we write (4.4) as

$$\begin{aligned} \int_{\bullet} \frac{\partial f}{\partial t} \xi &= \int_{\bullet} (\mathbf{y}^1 f) \cdot \nabla_{\mathbf{x}} \xi + \sum_{k=1}^{K-1} \frac{1}{\varepsilon_k} \int_{\bullet} ((\mathbf{y}^{k+1} - \alpha_k \mathbf{y}^k) f) \cdot \nabla_{\mathbf{y}^k} \xi \\ &\quad - \frac{1}{\varepsilon_K} \int_{\bullet} (\alpha_K \mathbf{y}^K f) \cdot \nabla_{\mathbf{y}^K} \xi + \frac{1}{2\varepsilon_K} \int_{\bullet} (G(W * \varrho)^2 f) \Delta_{\mathbf{y}^K} \xi. \end{aligned}$$

This is the weak formulation of the Fokker-Planck equation

$$\begin{aligned} \partial_t f &+ \mathbf{y}^1 \cdot \nabla_{\mathbf{x}} f + \sum_{k=1}^{K-1} \frac{1}{\varepsilon_k} \nabla_{\mathbf{y}^k} \cdot ((\mathbf{y}^{k+1} - \alpha_k \mathbf{y}^k) f) \\ &= \frac{1}{\varepsilon_K} \nabla_{\mathbf{y}^K} \cdot \left(\alpha_K \mathbf{y}^K f + \frac{G(W * \varrho)^2}{2} \nabla_{\mathbf{y}^K} f \right), \end{aligned} \tag{4.6}$$

with the particle number density $\varrho = \varrho(t, \mathbf{x})$ given by (4.3), i.e.,

$$\varrho(t, \mathbf{x}) = \int_{\mathbb{R}^{dK}} f(t, \mathbf{x}, \mathbf{y}) d\mathbf{y}. \tag{4.7}$$

5 Equilibrium patterns in the large population limit

Equilibrium patterns in the limit $N \rightarrow \infty$ of the discrete system (3.1) correspond to steady states of the Fokker-Planck equation (4.6)–(4.7). In Sections 6 and 7 we will present results of stochastic simulations of the discrete system where the positions of individuals, $\mathbf{x}_i = \mathbf{x}_i(t) \in \mathbb{R}^d$, $i \in [N]$, evolve in the d -dimensional cube

$$\Omega = (0, 1)^d \quad \text{with periodic boundary conditions.} \tag{5.1}$$

We therefore consider equation (4.6) posed on a torus Ω in the \mathbf{x} variable, while the \mathbf{y}^k variables, for $k \in [K]$, take their values in the full space \mathbb{R}^d . Since for any fixed $\mathbf{x} \in \Omega$ the function

$$\mathbf{y}^K \mapsto \exp \left(-\frac{\alpha_K |\mathbf{y}^K|^2}{G(W * \varrho(\mathbf{x}))^2} \right)$$

is an equilibrium for the Fokker-Planck operator on the right-hand side of equation (4.6), we have the following homogeneous (i.e., constant in \mathbf{x}) steady state solution of the Fokker-Planck equation (4.6)

$$f(t, \mathbf{x}, \mathbf{y}) = C \exp\left(-\frac{\alpha_K |\mathbf{y}^K|^2}{G(\bar{\varrho})^2}\right) \prod_{k=1}^{K-1} \delta(\mathbf{y}^k - \alpha_k^{-1} \mathbf{y}^{k+1}), \quad (5.2)$$

where $\bar{\varrho} > 0$ is the constant particle density in Ω , δ is the Dirac delta function on \mathbb{R}^d and $C > 0$ is the normalization constant. Using (4.7) and (5.2), we get

$$\bar{\varrho} = \int_{\mathbb{R}^{dK}} f(t, \mathbf{x}, \mathbf{y}) d\mathbf{y} = C \left(\frac{\pi}{\alpha_K}\right)^{d/2} G(\bar{\varrho})^d.$$

Consequently, the normalization constant C in equation (5.2) is given by

$$C = \left(\frac{\alpha_K}{\pi}\right)^{d/2} \frac{\bar{\varrho}}{G(\bar{\varrho})^d}.$$

In particular, equation (5.2) represents a steady state that is constant in \mathbf{x} , concentrated on the diagonals $\mathbf{y}^k = \alpha_k^{-1} \mathbf{y}^{k+1}$, for $k \in [K-1]$, and Gaussian distributed in the \mathbf{y}^K variable. From the point of view of the discrete individual-based model (3.1), this corresponds to the equilibrium where no aggregation takes place (constant particle density) and the internal variables \mathbf{y}^k , for $k \in [K]$, are subject to the Brownian motion. To characterise the steady states which are non-constant in the \mathbf{x} variable, we first write the momentum system corresponding to the Fokker-Planck equation (4.6). Integrating (4.6) with respect to all \mathbf{y}^k variables, we obtain

$$\partial_t \varrho + \nabla_{\mathbf{x}} \cdot \int_{\mathbb{R}^{dK}} \mathbf{y}^1 f d\mathbf{y} = 0. \quad (5.3)$$

We observe that constructing f radially symmetric in \mathbf{y}^1 (i.e., depending on \mathbf{y}^1 only through the modulus $|\mathbf{y}^1|$) annihilates the term $\int_{\mathbb{R}^{dK}} \mathbf{y}^1 f d\mathbf{y}$ and equation (5.3) reduces to $\partial_t \varrho = 0$. Multiplying equation (4.6) by \mathbf{y}^K and integrating with respect to all \mathbf{y}^k variables, we obtain

$$\partial_t (\varrho m^K) + \nabla_{\mathbf{x}} \cdot \int_{\mathbb{R}^{dK}} (\mathbf{y}^1 \otimes \mathbf{y}^K) f d\mathbf{y} = -\frac{1}{\varepsilon_K} \alpha_K \varrho m^K, \quad (5.4)$$

where we have denoted $\varrho m^K := \int_{\mathbb{R}^{dK}} \mathbf{y}^K f d\mathbf{y}$. From here we infer that in equilibrium we either have empty regions where $\varrho = 0$, or regions with positive density $\varrho > 0$, but zero moment $m^K = 0$. Moreover, let us observe that any equilibrium of (4.6) needs to be concentrated of the diagonals $\mathbf{y}^k = \alpha_k^{-1} \mathbf{y}^{k+1}$, for $k \in [K-1]$. Since we have imposed radial symmetry in \mathbf{y}^1 , this implies that the equilibrium is radially symmetric in all \mathbf{y}^k variables. Then an easy calculation reveals that

$$\left(\prod_{k=1}^{K-1} \alpha_k\right) \nabla_{\mathbf{x}} \cdot \int_{\mathbb{R}^{dK}} (\mathbf{y}^1 \otimes \mathbf{y}^K) f d\mathbf{y} = \nabla_{\mathbf{x}} \cdot \int_{\mathbb{R}^{dK}} (\mathbf{y}^K \otimes \mathbf{y}^K) f d\mathbf{y} = \frac{2}{d} \nabla_{\mathbf{x}} (\varrho e^K),$$

where we have denoted $\varrho e^K = \frac{1}{2} \int_{\mathbb{R}^{dK}} |\mathbf{y}^K|^2 f d\mathbf{y}$. We see that ϱe^K plays the role of pressure in the momentum equation (5.4) and in equilibrium we must have $\varrho e^K \equiv \text{constant}$ on Ω . Finally, multiplying equation (4.6) by $|\mathbf{y}^K|^2$ and integrating by parts, we obtain

$$\partial_t (\varrho e^K) + \nabla_{\mathbf{x}} \cdot \left(\frac{1}{2} \int_{\mathbb{R}^{dK}} |\mathbf{y}^K|^2 \mathbf{y}^1 f d\mathbf{y}\right) = -\frac{1}{\varepsilon_K} \alpha_K \varrho e^K + \frac{d}{2\varepsilon_K} G(W * \varrho)^2 \varrho.$$

We observe that in equilibrium we must have

$$\frac{d}{2} G(W * \varrho)^2 \varrho = \alpha_K \varrho e^K.$$

Due to the requirement that ϱe^K be constant in Ω , we finally arrive at the condition characterising the equilibrium density profiles, namely,

$$G(W * \varrho)^2 \varrho \equiv C_0 \quad \text{in } \Omega, \quad (5.5)$$

where the constant $C_0 > 0$ is determined by the initial mass. This allows for nonhomogeneous number density profiles $\varrho = \varrho(\mathbf{x})$, i.e., presence of clusters. To numerically compute such nonhomogeneous profiles $\varrho = \varrho(\mathbf{x})$ satisfying condition (5.5), we can solve equation (2.4) in the unstable regime, subject to an initial datum that is a perturbation of the constant steady state. We employ this strategy in the spatially one-dimensional setting, i.e., for $d = 1$ in our domain (5.1). The results are presented in Figure 2, where we use

$$G(s) := e^{-s} \quad \text{and} \quad W(x) := \frac{\chi_{[0,R]}(|x|)}{2R} \quad \text{with some } R > 0, \quad (5.6)$$

where $\chi_{[0,R]}$ denotes the characteristic function of the interval $[0, R]$, i.e., the kernel $W = W(x)$ corresponds to the sampling radius $R > 0$. Using the semi-implicit finite difference discretization for the space variable and first-order forward Euler method for the time variable, the steady states are plotted for $R \in \{0.1, 0.05, 0.025\}$ in Figure 2. Let us note that the steady states are not unique and the particular pattern produced by the simulation depends on the choice of the initial datum; for all three simulations presented in Figure 2 we used $\varrho(t = 0, x) = 1 + 10^{-1} \sin(2\pi x(2 - x))^2$. The same approach can be applied in higher-dimensional settings as well, see [8, Section 5.4] for the results with $d = 2$.

Finally, let us mention that in the formal limit as $R \rightarrow 0+$ in (5.6) we obtain $W(x) = \delta(x)$, the Dirac delta function in \mathbb{R}^d . Then $W * \varrho = \varrho$ and equation (2.4) becomes a nonlinear but local reaction-diffusion equation. However, it does not produce any nontrivial patterns, since the following alternative holds: either $G = G(s)$ is such that all homogeneous steady states are asymptotically stable, or equation (2.4) is ill-posed. We refer to [8, Section 4.4] for more details.

6 Spontaneous aggregation in one spatial dimension ($d = 1$)

As the dimensionality of the Fokker-Planck equation (4.6) becomes prohibitive for numerical simulations even with moderate values of K , we use stochastic simulations of the individual-based model given by equations (2.1) and (3.1) to systematically investigate the impact of the number of memory layers $K \geq 1$ on the clustering properties of the spontaneous aggregation model. We use ε_k and α_k , for $k \in [K]$, given by equations (3.3) and (3.7), respectively, where we choose $\alpha = 1$ in equation (3.7). The distance between agents is calculated over the torus, i.e., taking into account the periodic boundary conditions on Ω . We simulate $N = 400$ agents moving in the domain (5.1) for $d = 1$. The response function $G(s)$ and the interaction kernel $W(x)$ are given by (5.6) with the sampling radius $R = 1/40 = 0.025$.

We initialize the simulation by randomly generated agent positions $x_i \in \Omega$, for $i \in [N]$, using uniformly distributed initial positions in Ω . All internal variables are initialized as zeros, i.e., $y_i^k(0) = 0$ for all $i \in [N]$ and $k \in [K]$. We discretize equations (3.1) using the Euler-Maruyama scheme with timestep $\Delta t = 10^{-3}$. We calculate 100 stochastic realizations of the individual-based model in the time interval $[0, 10^3]$, i.e., we calculate the time evolution over 10^6 timesteps for

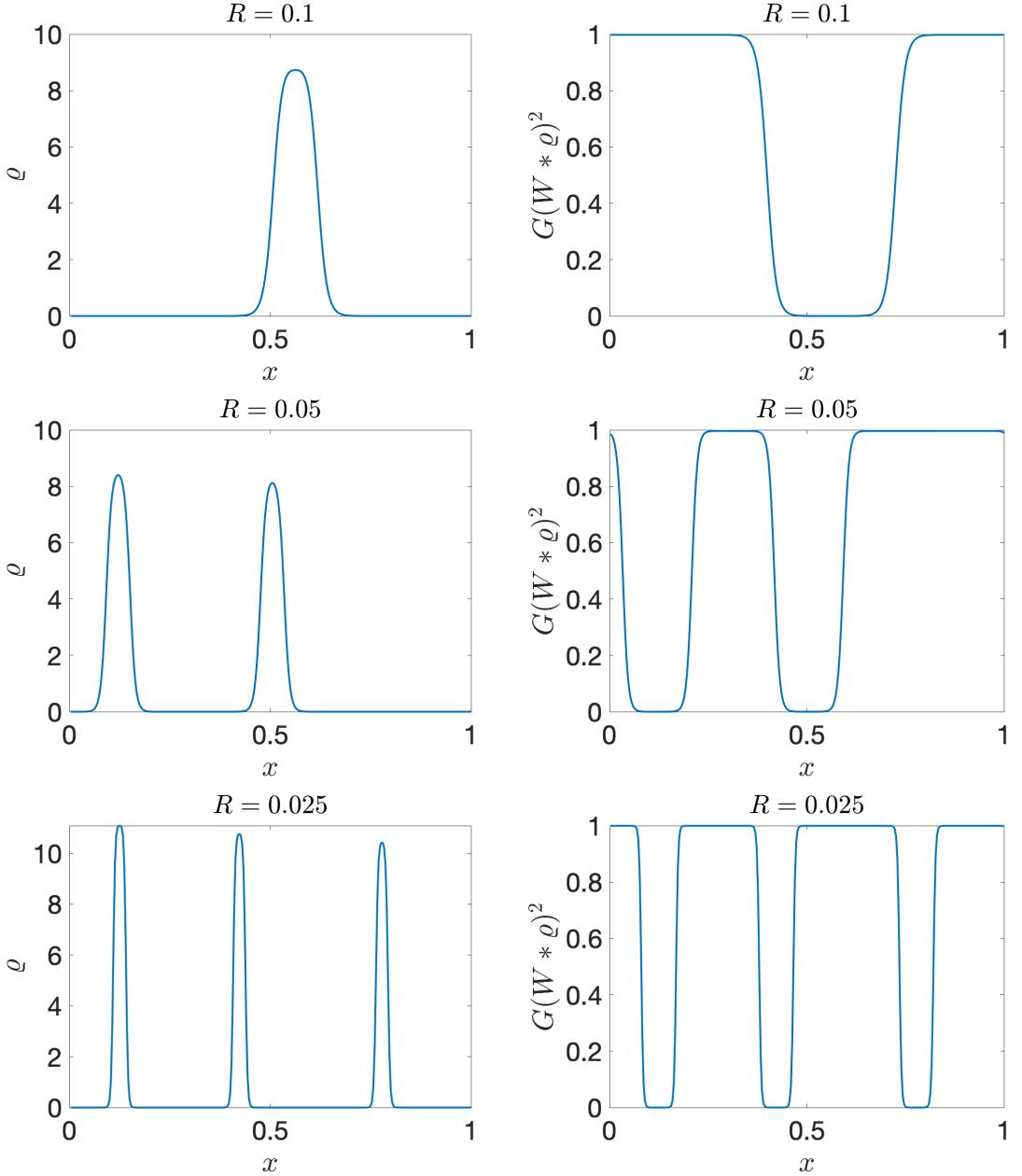


Figure 2: One-dimensional equilibrium profiles satisfying the condition (5.5), obtained by solving equation (2.4) in domain (5.1) for $d = 1$, subject to the initial datum $\rho(t = 0, x) = 1 + 10^{-1} \sin(2\pi x(2 - x))^2$, until a steady state is reached. We use $G(s)$ and $W(x)$ given by (5.6) with $R = 0.1$ (top), $R = 0.05$ (middle) and $R = 0.025$ (bottom). The left panels depict the steady state density $\rho = \rho(x)$, while the right panels visualize the function $G(W * \rho)^2$.

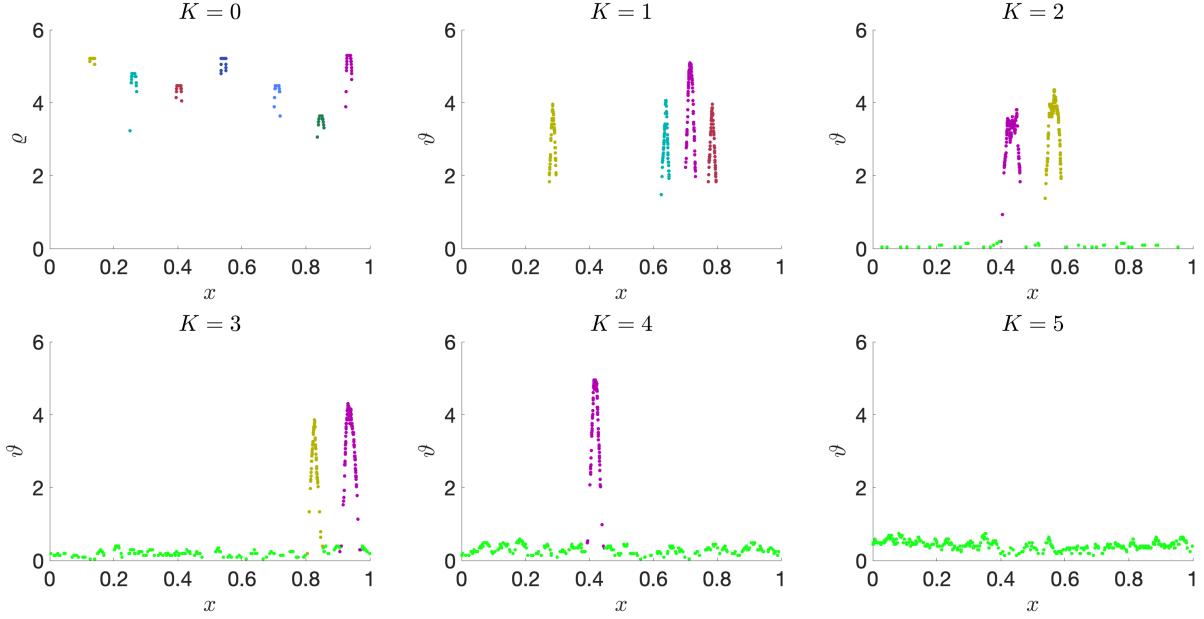


Figure 3: Simulations of the individual-based model (2.1) and (3.1) for $K \in \{1, \dots, 5\}$, and the model (2.1)–(2.3) without memory ($K = 0$). We used $N = 400$ agents moving in the domain Ω given by (5.1) with $d = 1$. $G(s)$ and $W(\mathbf{x})$ are given by (5.6) with $R = 0.025$. For each value of $K \in \{0, 1, \dots, 5\}$, the plots capture the particle positions at the final time $t = 10^3$ (horizontal axis) and the perceived density of their neighbours ϑ_i given by (2.1) (vertical axis). The clusters, differentiated by colour, are identified using the DBSCAN method with parameters `epsilon` = 0.025 and `MinPoints` = 20. The light green points are outliers, i.e., particles not belonging to any cluster.

each realization. Initially, clusters emerge from random fluctuations in the particle distribution. Subsequently, small clusters dissociate and a dynamic equilibrium is established between clustered particles and freely moving outliers. This equilibrium typically forms on a relatively fast timescale, e.g., at the order of the first 10^5 timesteps (out of the total 10^6 timesteps carried out in the simulation). Then, the number of clusters remains constant until the end of the simulation. The location, shape, and size of the clusters fluctuate only minimally. We thus observe the emergence of well-defined quasi-stationary states. We record the particle positions at the final time $t = 10^3$ and use these for evaluating clustering properties of the model.

For identification of the clusters we use the MATLAB implementation of the Density-based spatial clustering (DBSCAN) method [30]. As we expect the clusters to be of size comparable to the sampling radius $R = 0.025$, we set the parameter `epsilon`, specifying the radius of a neighbourhood with respect to some point of the DBSCAN method, to `epsilon` = 0.025. Moreover, based on experimentation, we found that the choice of the second parameter of DBSCAN, `minPts` = 20, leads to the best results in identification of clusters. Examples of the results, recorded at the final timestep of the simulations with $K \in \{1, 2, 3, 4, 5\}$, are plotted in Figure 3. For comparison, we also simulate the system (2.1)–(2.3), i.e., spontaneous aggregation without memory; we refer to the corresponding results by $K = 0$.

In Figure 4, we provide statistics of the clustering behaviour over 100 realizations of the individual-based stochastic model given by equations (2.1) and (3.1), performed for each value $K \in \{1, 2, \dots, 6\}$, together with the results obtained by the system (2.1)–(2.3) (referred as the $K = 0$ case). In Figure 4(a), we plot the number of clusters identified in the final timestep of the

simulation by the DBSCAN method. The solid line represents the average over the 100 realizations, while the blue ‘error bars’ indicate the minimum and maximum values. Here, we clearly observe a tendency toward the formation of fewer clusters as K increases. Figure 4(b) shows the minimum, average and maximum cluster sizes observed over the course of 100 stochastic realizations. For $K \in \{0, 1, 2\}$ the size of the clusters increases with K , and decreases for $K > 3$. Figure 4(c) shows the average (solid line) and minimum/maximum (error bars) number of outliers, i.e., particles that do not belong to any cluster. The number of outliers increases from almost zero for $K \in \{0, 1\}$, up to almost all particles being outliers for $K \in \{5, 6\}$; c.f. the corresponding panels in Figure 3. Finally, the percentages of the outcomes (out of the 100 stochastic simulations) that did not produce any clusters are plotted in Figure 4(d). We observe that for $K \in \{0, 1, 2, 3, 4\}$ the percentage is zero (clusters have always been formed). For $K = 5$ the percentage increases steeply and is almost 100% for $K = 6$.

In summary, increasing memory length disrupts spontaneous aggregation. In the next section, we present an analogous set of simulations in a two-dimensional spatial setting. There, we also provide statistical evidence that memory weakens the particles’ responsivity to environmental stimuli. We argue that this reduction in responsivity accounts for the decreased tendency to cluster observed at higher values of K .

7 Spontaneous aggregation in two spatial dimensions ($d = 2$)

In this section, we present results of stochastic simulations of the individual-based model given by equations (2.1) and (3.1) for $d = 2$ (spatially two-dimensional setting), where $\mathbf{x}_i = \mathbf{x}_i(t) \in \Omega$ with Ω given by (5.1) for $d = 2$. The internal variables $\mathbf{y}_i^k = \mathbf{y}_i^k(t)$ evolve in the full space \mathbb{R}^2 . As in Section 6, we set ε_k and α_k , for $k \in [K]$, given by equations (3.3) and (3.7), respectively, where we choose $\alpha = 1$ in equation (3.7). We simulate $N = 400$ particles with the response function $G(s)$ and the interaction kernel $W(\mathbf{x})$ given by

$$G(s) := e^{-s} \quad \text{and} \quad W(\mathbf{x}) := \frac{\chi_{[0,R]}(|\mathbf{x}|)}{\pi R^2}, \quad (7.1)$$

where we choose the sampling radius $R = 1/20 = 0.05$. We again discretize (3.1) using the Euler-Maruyama scheme with timestep $\Delta t = 10^{-3}$. We initialize the simulation by generating initial positions $\mathbf{x}_i \in \Omega$, for $i \in [N]$, using a uniform distribution in Ω , while all internal variables are initially equal to zero, i.e., $\mathbf{y}_i^k(0) = \mathbf{0}$ for all $i \in [N]$ and $k \in [K]$. We again calculate 100 stochastic realizations, now for $K \in \{1, 2, \dots, 8\}$, each time performing 10^7 timesteps. We record the particle locations at the last timestep ($t = 10^4$) and use these for evaluating clustering properties of the model. For identification of the clusters we again use the MATLAB implementation of the DBSCAN method [30] with its parameter chosen as `epsilon` = 0.05 (same as the sampling radius) and `minPts` = 12.

Examples of the results, recorded at the final timestep of the simulations with $K \in \{1, 2, \dots, 8\}$, are plotted in Figure 5. The clusters are indexed by positive natural numbers, while agents indexed by -1 are classified as outliers, i.e., not belonging to any cluster. We observe the tendency to produce a smaller number of clusters with increasing K . The size of clusters appears to increase until $K = 4$. For larger values of K the number of outliers increases. For $K = 8$ no clusters are formed in the presented stochastic realization in Figure 5, i.e., all particles have been classified as outliers.

In Figure 6, we provide statistics of the clustering behaviour over the sample of 100 realizations of the individual-based model given by equations (2.1) and (3.1) for $d = 2$, performed for each

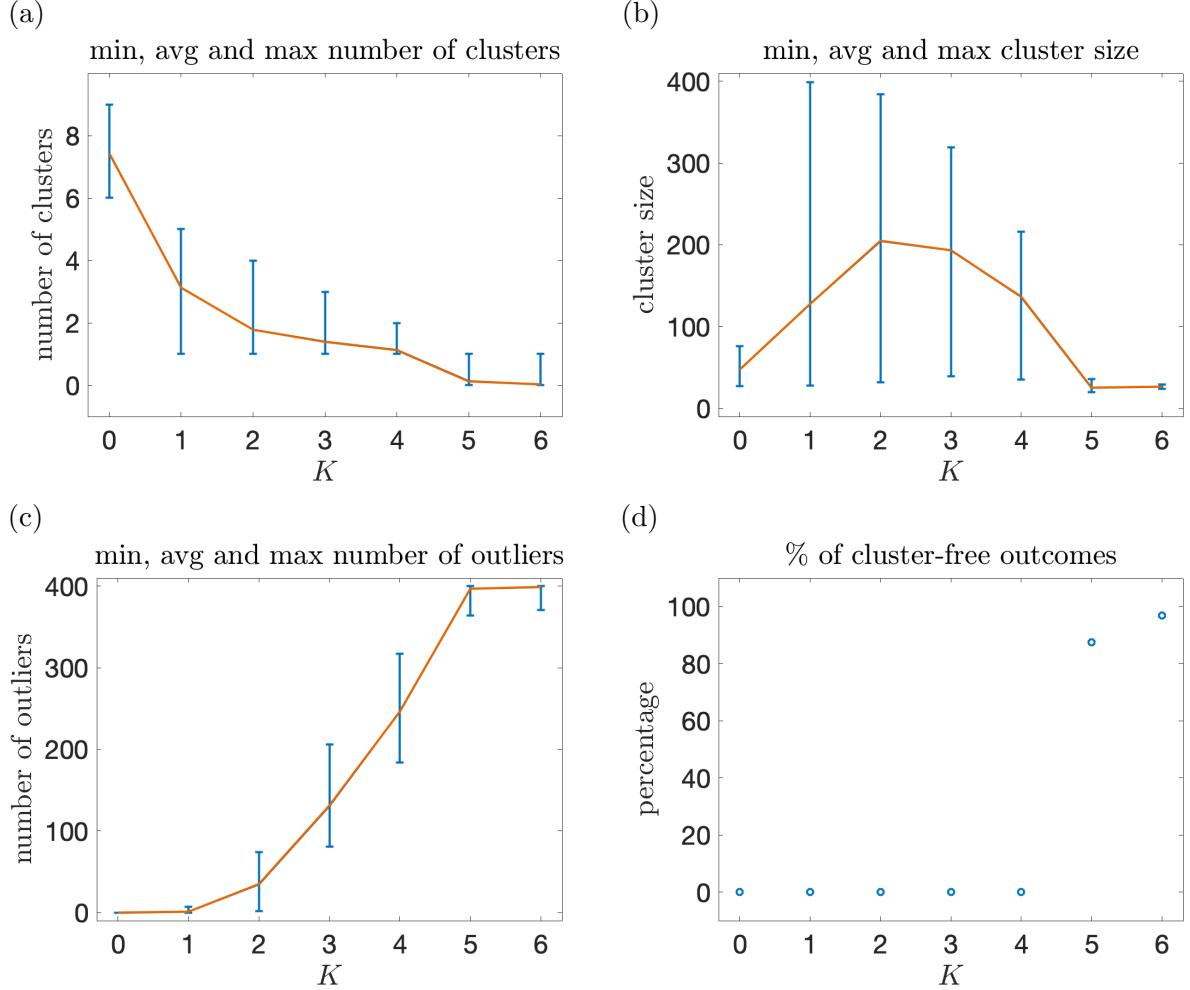


Figure 4: Statistics of the clustering behaviour over 100 realizations of the individual-based model given by equations (2.1) and (3.1) for dimension $d = 1$, $N = 400$ agents, $G(s)$ and $W(x)$ given by (5.6) with $R = 0.025$ and $K \in \{1, 2, \dots, 6\}$. The case $K = 0$ (no memory) refers to stochastic simulation of the system (2.1)–(2.3). Other parameters are the same as in Figure 3.

- (a) average (orange solid line), minimum and maximum (indicated by blue error bars) number of clusters identified in the final timestep of the simulations at time $t = 10^3$,
- (b) average (orange solid line), minimum and maximum (blue error bars) cluster sizes,
- (c) average (orange solid line), minimum and maximum (blue error bars) number of outliers, i.e., particles that do not belong to any cluster,
- (d) percentage of simulation outcomes (out of the 100 runs) that did not produce any clusters.

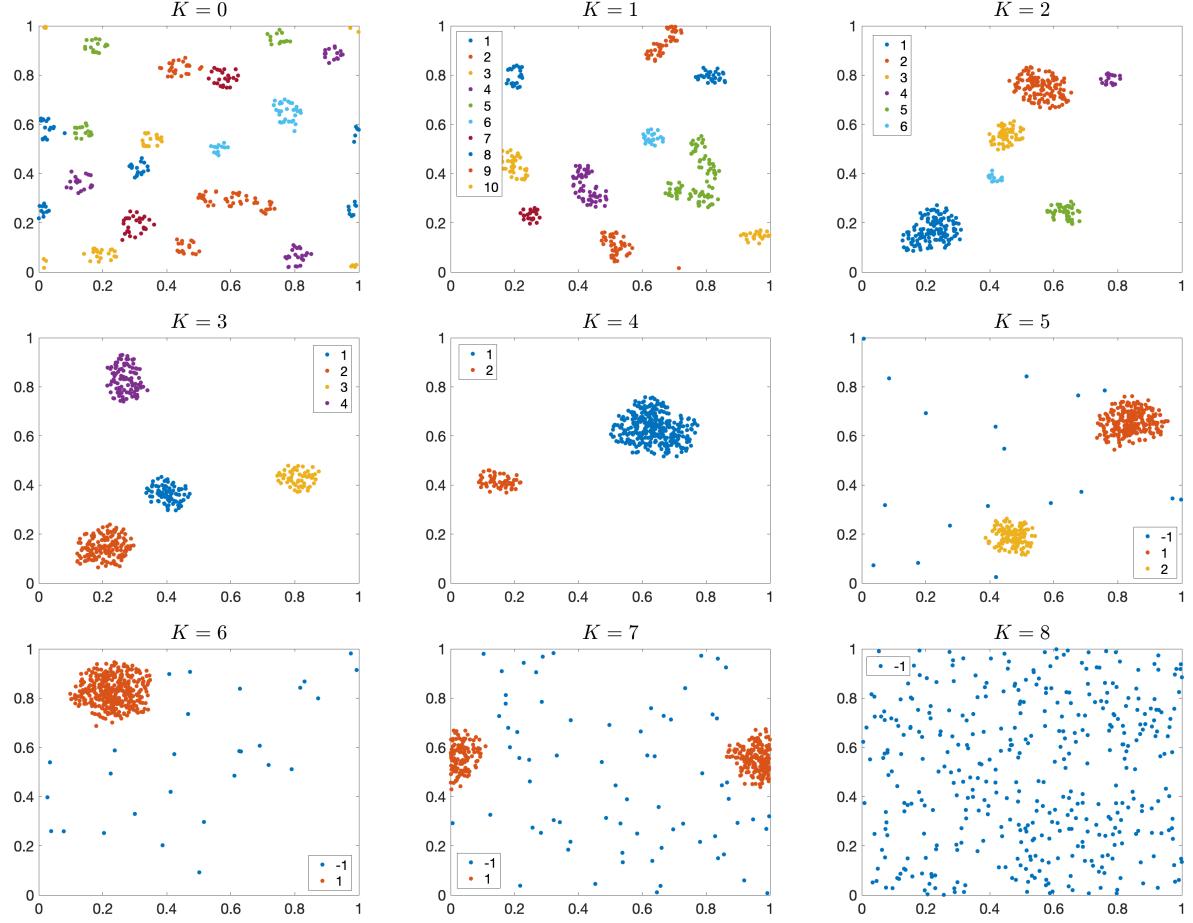


Figure 5: Simulations of the individual-based model (2.1) and (3.1) for $K \in \{1, \dots, 8\}$, and the model (2.1)–(2.3) without memory ($K = 0$). We used $N = 400$ agents moving in the domain Ω given by (5.1) with $d = 2$. Functions $G(s)$ and $W(\mathbf{x})$ are given by (5.6) with $R = 0.05$. For each value of $K \in \{0, 2, \dots, 8\}$, the plots capture the particle positions at the final time $t = 10^4$. The clusters, differentiated by colour, are identified using the DBSCAN method with parameters $\text{epsilon} = 0.05$ and $\text{minPts} = 12$. The points indexed with -1 are outliers, i.e., not belonging to any cluster.

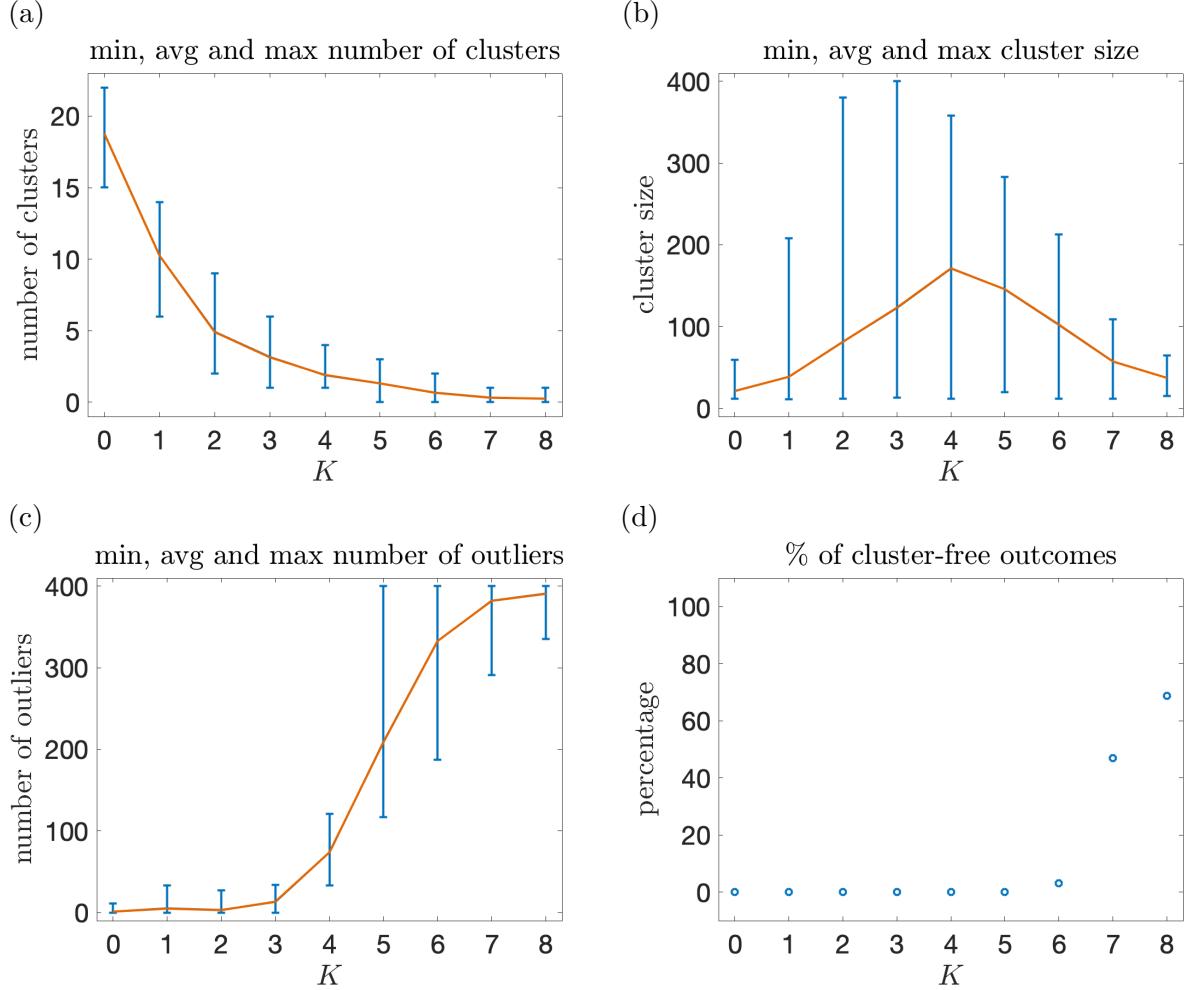


Figure 6: *Statistics of the clustering behaviour over 100 realizations of the individual-based model given by equations (2.1) and (3.1) for dimension $d = 2$, $N = 400$ particles, $G(s)$ and $W(\mathbf{x})$ given by (7.1) with $R = 0.05$ and $K \in \{1, 2, \dots, 8\}$. The case $K = 0$ (no memory) refers to stochastic simulation of the system (2.1)–(2.3). Other parameters are the same as in Figure 5.*

- (a) average (orange solid line), minimum and maximum (indicated by blue error bars) number of clusters identified in the final timestep of the simulations at time $t = 10^3$,
- (b) average (orange solid line), minimum and maximum (blue bars) cluster sizes,
- (c) average (orange solid line), minimum and maximum (blue bars) number of outliers, i.e., particles that do not belong to any cluster,
- (d) percentage of simulation outcomes (out of the 100 runs) that did not produce any clusters.

$K \in \{1, 2, \dots, 8\}$. Again, we also simulate the memoryless system (2.1)–(2.3) and refer to the results by $K = 0$. The statistical quantities plotted in the four panels of Figure 6 are the same as in the one-dimensional case in Figure 4. In particular, Figure 6(a) shows the average, minimum and maximum number of clusters formed obtained over 100 stochastic realizations. We again observe the tendency to formation of fewer clusters for higher values of K . For $K \in \{1, 2, 3, 4\}$ the size of the clusters increases with K , as we observe in Figure 6(b), and decreases for $K > 4$. Figure 6(c) shows the average (orange solid line) and minimum/maximum (blue error bars) number of outliers, i.e., particles that do not belong to any cluster. We can again observe relatively low number of outliers for $K \leq 4$ and high number for $K \geq 5$. Finally, Figure 6(d) shows the percentage of the outcomes out of the 100 simulations that have not produced any clusters. We observe that for $K \in \{1, 2, 3, 4, 5\}$ the percentage is zero (clusters have always been formed). For larger values of K the percentage increases steeply. For $K = 8$ no clusters have been formed in 75 out of the 100 simulations. Based on the statistics gathered in Figure 6, we may identify three regimes:

- (a) *Short memory regime* ($K = 1$), where a high number of clusters (10 on average) is formed, typically consisting of 40 agents. There are no or almost no outliers.
- (b) *Moderate memory regime* ($K = 4$), where typically a few larger clusters form, containing a significant proportion of agents except for the outliers.
- (c) *Long memory regime* ($K = 8$) where most of the simulations do not produce any clusters, and even if a cluster is formed, the majority of agents exists as outliers.

To gain further understanding of this behaviour, we have collected the statistics of the values of the internal variables \mathbf{y}_i^1 , for $i \in [N]$, recorded during the temporal evolution of the individual-based model in the considered 100 stochastic realizations in Figure 7. More precisely, the forward Euler discretization of the first equation of the SDE system (3.1) reads

$$\mathbf{x}_i(t + \Delta t) = \mathbf{x}_i(t) + \mathbf{y}_i^1(t) \Delta t \quad \text{for } i \in [N],$$

i.e., in each timestep of the discrete simulation the particle locations in the physical space are updated by $\mathbf{y}_i^1(t) \Delta t$. In Figure 7, we plot the statistics of the magnitudes $|\mathbf{y}_i^1(t)|$ in dependence of the values of $G(\vartheta_i(t))$ with the perceived densities $\vartheta_i(t)$ given by equation (2.1). In particular, for each fixed $K \in [8]$, in each timestep of each of the 100 stochastic realizations, we record the pair $(G(\vartheta_i(t)), |\mathbf{y}_i^1(t)|)$ for all $i \in [N]$. We then plot the mean and standard deviation of the values of $|\mathbf{y}_i^1|$ against the values of $G(\vartheta_i)$, separately for each memory regime in panels (a), (b) and (c) of Figure 7. Figure 7(a) corresponds to the short memory regime, Figure 7(b) presents the moderate memory regime and Figure 7(c) corresponds to the long memory regime. We observe that $|\mathbf{y}_i^1|$ increases with $G(\vartheta_i)$, which corresponds to the fundamental modeling assumption that the particles' mobility decreases with perceived density (recall that the response function G is in general a decreasing function; we use $G(s) = e^{-s}$ in equation (7.1)). However, this effect weakens as K increases, and is nearly absent at $K = 8$, see Figure 7(c). We therefore conclude that the presence of memory systematically inhibits the particles' responsivity to environmental stimuli, specifically the perceived density of their neighbours.

8 Discussion

In this paper, we have shown that short-term memory enhances and long-term memory inhibits spontaneous particle aggregation. Our investigation has been based on the (first-order) spontaneous aggregation model without memory which has been previously investigated in the literature [8]. Its

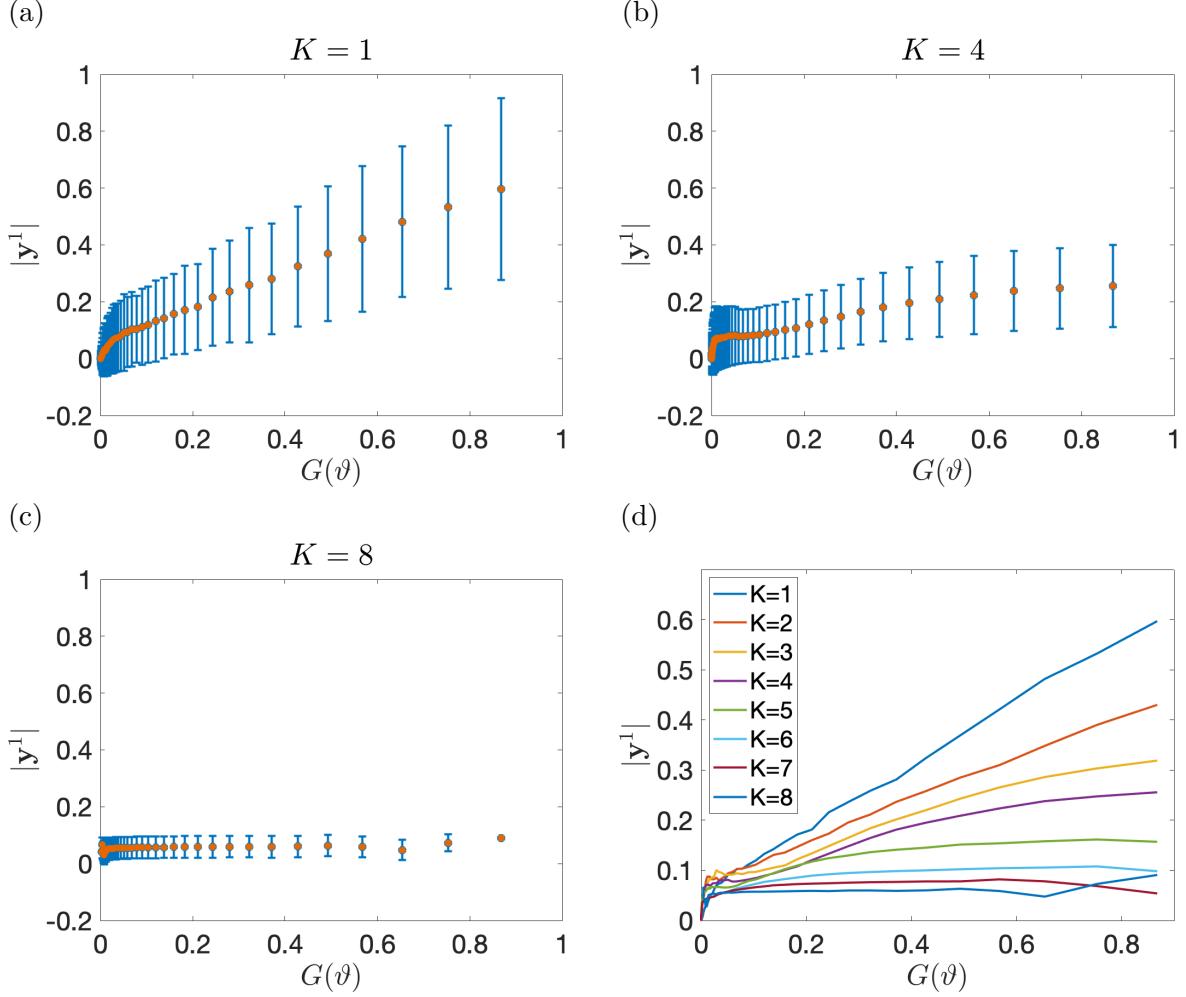


Figure 7: Statistics of the magnitude of the internal variable $|\mathbf{y}_i^1|$ versus the response function value $G(\vartheta_i)$, aggregated over all agents and timesteps in 100 simulations of the individual-based model given by equations (2.1) and (3.1) in $d = 2$ dimensions with $N = 400$ particles, for:

- (a) $K = 1$ (short memory regime),
- (b) $K = 4$ (moderate memory regime),
- (c) $K = 8$ (long memory regime).

The orange circles represent the average value of $|\mathbf{y}_i^1|$ corresponding to each $G(\vartheta_i)$, while the error bars indicate the standard deviations. The panel (d) presents the average values of $|\mathbf{y}_i^1|$ corresponding to each $G(\vartheta_i)$ for all values $K \in \{1, 2, \dots, 8\}$.

main properties are summarized in Section 2. The memory has been added into this model in Section 3 by introducing a chain of K internal variables that allow the agents to ‘remember’ the densities they encountered in the past. If $K = 1$, then our individual-based model is equivalent to what is called ‘the second-order model’ in reference [8], where the internal variable represents the agent’s velocity. Considering $K > 1$, the model introduces additional layers of memory described by K internal variables. Additional internal variables can be used to better fit the properties of relatively complex (high-dimensional) models of interacting particle systems, while keeping the number of degrees of the system (internal variables) relatively small [21, 22]. For simplicity, our transfer of information between layers of memory is linear, but nonlinear functions can also be introduced to better fit the properties of some systems [22, 33, 23].

In Sections 6 and 7, we have reported the results of systematic stochastic simulations of the individual-based model in the spatially one-dimensional and two-dimensional settings, respectively. We have shown that memory inhibits the particles’ responsivity to environmental stimuli. Our results show that the introduction of memory leads, in general, to the formation of a smaller number of larger clusters. This trend is observed until $K = 3$ in the one-dimensional setting and until $K = 4$ in two spatial dimensions. When the number of layers K is increased further, i.e., as the memory becomes ‘longer’, its effect starts to be disruptive. This is manifested by the increasing proportion of ‘outliers’, which are the particles that are not part of any cluster. Also, the percentage of stochastic simulations where no clusters are formed during the observation time increases with increasing K . We therefore conclude that short-term or medium-term memory has a coarsening effect on spontaneous aggregation, while long-term memory disrupts it.

The presence of memory impacts the clustering properties of the model - in particular, the number and size of clusters being formed, or even the very ability of the model to produce clusters, starting from initially randomly distributed particles in the physical space. In one spatial dimension, we have observed a sharp transition between $K \leq 4$, when clusters are always formed, and $K \geq 5$, when clusters are almost never formed during the simulation period. A similar, although less sharp transition, takes place in two-dimensional models between $K \leq 6$ and $K \geq 7$.

We have performed systematic numerical simulations to evaluate the impact of the number of internal variables (or ‘memory layers’) K on the dynamics of the many particle system, because (exact) analytical mathematical results to study these properties are not available and the development of the corresponding mathematical theories includes a number of open questions. Some progress in this direction can be made as shown in Section 4, where we have derived the formal macroscopic limit of the system as the number of agents tends to infinity, under the usual molecular chaos assumption [8, 25]. The limit is described by a Fokker-Planck equation, and we have characterized its steady states to gain an insight into the patterns (clusters) formed by the system.

Many organisms benefit from learning to adapt their behaviour to the distribution of resources and their learning depends on the memory capacity of each individual [32]. In our investigation, the number of memory layers, K , has been fixed, but it could be made variable for each agent. If the evolutionary objective is to develop large clusters in two-dimensional setting, then the choice $K = 4$ can be considered ‘optimal’. Our memory model with K layers could also be further extended by considering nonlinear activation functions to model a more complicated neural network [33] and the applicability of the model could be further enhanced by introducing different modes of learning during the individual-based simulations [43].

In our computational simulations, we have used exponentially decaying function $G(s) = e^{-s}$ in equations (5.6) and (7.1) in order for clusters to be formed on effectively ‘intermediate’ timescales of the order 10^3 . Since all of the simulated particles are subject to the Brownian motion (including the particles in the clusters), the clusters are also able to slowly move (diffuse) on much larger timescales. However, as we are interested in biological applications, such long time behaviour of

the model can be irrelevant, because other processes (e.g birth or death of individuals) could also happen on long timescales. We have therefore run the simulation until clusters emerge (starting from the initial random distribution of particles) and become ‘quasi-stationary’ in the sense that they do not move (on this timescale) and are in dynamic equilibrium with the outliers. For a study of the clustering behaviour on long timescales, see, e.g., [9].

Declarations

Authorship and Contributorship

R.E. and J.H. contributed equally to the conception and writing of this manuscript. Both authors have read and approved the final manuscript. In addition, the following contributions occurred: J.H. developed the computer code, performed the numerical simulations, and prepared all figures.

Conflict of interests

The authors declare no conflict of interests.

Data & Code Availability

In compliance with EPSRC’s open access initiative, the data and code used in this paper is available from <http://dx.doi.org/xxx/xxx> (doi will be created/inserted at the time of acceptance).

Ethics

This study did not involve human participants, animals, or sensitive data requiring ethical approval or consent to participate.

Funding

This work was supported by the Engineering and Physical Sciences Research Council, grant number EP/V047469/1, awarded to Radek Erban.

Acknowledgements

The authors thank to Martin Burger (DESY and Universität Hamburg) for stimulating discussions that contributed to the preparation of this paper.

References

- [1] H. Ahn and M. Kang. Emergent dynamics of various Cucker–Smale type models with a fractional derivative. *Mathematical Biosciences and Engineering*, 20(10):17949–17985, 2023.
- [2] U. Alon, M Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature (London)*, 397:168–171, 1999.
- [3] H. Berg and D. Brown. Chemotaxis in *Escherichia coli* analysed by three-dimensional tracking. *Nature*, 239:500–504, 1972.

- [4] H. C. Berg. Bacterial microprocessing. *Cold Spring Harbor Symp. Quantit. Biol.*, 55:539–545, 1990.
- [5] A. Bernoff and C. Topaz. Nonlocal aggregation models: A primer of swarm equilibria. *SIAM Review*, 55(4):709–747, 2013.
- [6] G. Boschi, C. Cammarota, and R. Kühn. Opinion dynamics with emergent collective memory: The impact of a long and heterogeneous news history. *Physica A: Statistical Mechanics and its Applications*, 569:125799, 2021.
- [7] C. Brangwynne, C. Eckmann, D. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Julicher, and Hyman A. Germline p granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, 324:1729–1732, 2009.
- [8] M. Burger, J. Haskovec, and M.-T. Wolfram. Individual-based and mean-field modelling of direct aggregation. *Physica D - Nonlinear Phenomena*, 260:145–158, 2012.
- [9] J.A. Carrillo, K. Craig, and Y. Yao. Aggregation-diffusion equations: Dynamics, asymptotics, and singular limits. In N. Bellomo, P. Degond, and E. Tadmor, editors, *Active Particles, Volume 2: Advances in Theory, Models, and Applications*, Modeling and Simulation in Science, Engineering and Technology, pages 65–108. Birkhäuser, Cham, Switzerland, 2019.
- [10] J.A. Carrillo, R. Eftimie, and F. Hoffmann. Non-local kinetic and macroscopic models for self-organised animal aggregations. *Kinetic and Related Models*, 8(3):413–441, 2015.
- [11] P. Cattiaux, F. Delebecque, and L. Pédèches. Stochastic cucker–smale models: Old and new. *The Annals of Applied Probability*, 28(5):3239–3286, 2018.
- [12] A. Celani and M. Vergassola. Bacterial strategies for chemotaxis response. *Proc Natl Acad Sci USA*, 107(4):1391–6, 2010.
- [13] C. Cercignani. *The Boltzmann Equation and Its Applications*, volume 67 of *Applied Mathematical Sciences*. Springer, 1988.
- [14] Z. Chen and Y. Zheng. Persistent and responsive collective motion with adaptive time delay. *Science Advances*, 10(14):eadk3914, 2024.
- [15] Y.-P. Choi and J. Haskovec. Hydrodynamic Cucker–Smale model with normalized communication weights and time delay. *SIAM Journal on Mathematical Analysis*, 51(3):2660–2685, 2019.
- [16] Y.-P. Choi, A. Paolucci, and C. Pignotti. Consensus of the Hegselmann–Krause opinion formation model with time delay. *Mathematical Methods in the Applied Sciences*, 44(6):4560–4579, 2021.
- [17] C. Cicolani, B. Ouahab, and C. Pignotti. Opinion dynamics under common influencer assumption or leadership control. arXiv:2407.16901, Jul 2024.
- [18] I. Couzin, J. Krause, R. James, G. Ruxton, and N. Franks. Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 218(1):1–11, 2002.
- [19] F. Cucker and S. Smale. Emergent behaviour in flocks. *IEEE Transactions on Automatic Control*, 52(5):852–862, 2007.

- [20] M.R. D’Orsogna, Y.C. Chuang, A. Bertozzi, and L. Chayes. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. *Physical Review Letters*, 96(10):104302, 2006.
- [21] R. Erban. Coupling all-atom molecular dynamics simulations of ions in water with brownian dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2186):20150556, 2016.
- [22] R. Erban. Coarse-graining molecular dynamics: stochastic models with non-Gaussian force distributions. *Journal of Mathematical Biology*, 80(1):457–479, 2020.
- [23] R. Erban. Neural networks for learning macroscopic chemotactic sensitivity from microscopic models. *arXiv preprint arXiv:2509.12131*, 2025. Submitted to SIAM Journal on Life Sciences.
- [24] R. Erban and S.J. Chapman. *Stochastic Modelling of Reaction-Diffusion Processes*. Cambridge Texts in Applied Mathematics, Cambridge University Press, 2020.
- [25] R. Erban and J. Haskovec. From individual to collective behaviour of coupled velocity jump processes: A locust example. *Kinetic and Related Models*, 5(4):817–842, 2012.
- [26] R. Erban, J. Haskovec, and Y. Sun. On Cucker-Smale model with noise and delay. *SIAM Journal on Applied Mathematics*, 76(4):1535–1557, 2016.
- [27] R. Erban and H. Othmer. From individual to collective behaviour in bacterial chemotaxis. *SIAM Journal on Applied Mathematics*, 65(2):361–391, 2004.
- [28] R. Erban and H. Othmer. From signal transduction to spatial pattern formation in *E. coli*: A paradigm for multi-scale modeling in biology. *Multiscale Modeling and Simulation*, 3(2):362–394, 2005.
- [29] R. Erban and H. Othmer. Taxis equations for amoeboid cells. *Journal of Mathematical Biology*, 54(6):847–885, 2007.
- [30] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, and U. Fayyad, editors, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231. AAAI Press, 1996.
- [31] W. Fagan, F. McBride, and L. Koralov. Reinforced diffusions as models of memory-mediated animal movement. *Journal of the Royal Society Interface*, 20(200):20220700, 2023.
- [32] A. Falcón-Cortés, D. Boyer, M. Aldana, and G. Ramos-Fernández. Lévy movements and a slowly decaying memory allow efficient collective learning in groups of interacting foragers. *PLOS Computational Biology*, 19(10):1–19, 2023.
- [33] K. Gurney. *An introduction to neural networks*. CRC press, 2018.
- [34] J. Haskovec. Flocking dynamics and mean-field limit in the Cucker–Smale-type model with topological interactions. *Physica D: Nonlinear Phenomena*, 261:42–51, 2013.
- [35] J. Haskovec. A simple proof of asymptotic consensus in the Hegselmann–Krause and Cucker–Smale models with normalization and delay. *SIAM Journal on Applied Dynamical Systems*, 20(1):130–148, 2021.

- [36] J. Haskovec and D. Oelz. A free boundary problem for aggregation by short range sensing and differentiated diffusion. *DCDS-B*, 20:1461–1480, 2015.
- [37] J. Haskovec and C. Schmeiser. Stochastic particle approximation for measure valued solutions of the 2D Keller–Segel system. *Journal of Statistical Physics*, 135(1):133–151, 2009.
- [38] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3):1–24, 2002.
- [39] P.E. Jabin and S. Motsch. Clustering and asymptotic behavior in opinion formation. *Journal of Differential Equations*, 257:4165–4187, 2014.
- [40] R. Jeanson, C. Rivault, J.-L. Deneubourg, S. Blanco, R. Fournier, C. Jost, and G. Theraulaz. Self-organised aggregation in cockroaches. *Anim. Behav.*, 69:169–180, 2005.
- [41] E. Keller and L. Segel. Initiation of slime mold aggregation viewed as an instability. *Journal of Theoretical Biology*, 26(3):399–415, 1970.
- [42] D. Kim, P. Thompson, D. Wolfson, J. Merkle, L. Oliveira-Santos, J. Forester, T. Avgar, M. Lewis, and J. Fieberg. Identifying signals of memory from observations of animal movements. *Movement Ecology*, 12(1):72, 2024.
- [43] M. Lewis, W. Fagan, M. Auger-Méthé, J. Frair, J. Fryxell, C. Gros, E. Gurarie, S. Healy, and J. Merkle. Learning and animal movement. *Frontiers in Ecology and Evolution*, 9:681704, 2021.
- [44] C. López, E. Colombo, E. Hernández-García, and R. Martínez-García. Spatial competition and repulsion: pattern formation and the role of movement. In L. Giuggioli and P. Maini, editors, *The Mathematics of Movement: an Interdisciplinary Approach to Mutual Challenges in Animal Ecology and Cell Biology*. Springer, 2025.
- [45] X. Mao. *Stochastic Differential Equations and Applications*. Elsevier, Amsterdam, Netherlands, 2007.
- [46] A. Mogilner and L. Edelstein-Keshet. A non-local model for a swarm. *Journal of Mathematical Biology*, 38(6):534–570, 1999.
- [47] J. Nauta, Y. Khaluf, and P. Simoens. Hybrid foraging in patchy environments using spatial memory. *Journal of the Royal Society Interface*, 17(166):20200026, 2020.
- [48] R. Olfati-Saber. Flocking for multi-agent dynamic systems: Algorithms and theory. *IEEE Transactions on Automatic Control*, 51(3):401–420, 2006.
- [49] M. Rosvall, A.V. Esquivel, A. Lancichinetti, J.D. West, and R. Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, 5, 2014.
- [50] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C.J. Tessone, and F. Schweitzer. Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nature Communications*, 5:5024, 2014.
- [51] J. Shen. Cucker–Smale flocking under hierarchical leadership. *SIAM Journal on Applied Mathematics*, 68(3):694–719, 2007.

- [52] P. Spiro, J Parkinson, and H. Othmer. A model of excitation and adaptation in bacterial chemotaxis. *Proceedings of the National Academy of Sciences USA*, 94:7263–7268, 1997.
- [53] D. Sumpter. The principles of collective animal behaviour. *Philosophical Transactions of The Royal Society B*, 361:5–22, 2006.
- [54] D. Sumpter. *Collective Animal Behavior*. Princeton University Press, 2010.
- [55] Y. Sun, W. Lin, and R. Erban. Time delay can facilitate coherence in self-driven interacting particle systems. *Physical Review E*, 90(6):062708, 2014.
- [56] J. Taylor-King, B. Franz, C. Yates, and R. Erban. Mathematical modelling of turning delays in swarm robotics. *IMA Journal of Applied Mathematics*, 80(5):1454–1474, 2015.
- [57] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6):1226–1229, 1995.
- [58] T. Vicsek and A. Zafeiris. Collective motion. *Physics Reports*, 517(3–4):71–140, 2012.
- [59] K. Vogtt, G. Beaucage, K. Rishi, H. Jiang, and A. Mulderig. Hierarchical approach to aggregate equilibria. *Phys. Rev. Res.*, 1:033081, 2019.
- [60] O.E. Williams, L. Lacasa, A.P. Millán, J. Buldú, and M.A. Porter. The shape of memory in temporal networks. *Nature Communications*, 13:499, 2022.